

Real-Time Big-Data Warehouse a Next-Generation Enterprise Data Warehouse and Analysis Framework

Authors : Abbas Raza Ali

Abstract : Big Data technology is gradually becoming a dire need of large enterprises. These enterprises are generating massively large amount of off-line and streaming data in both structured and unstructured formats on daily basis. It is a challenging task to effectively extract useful insights from the large scale datasets, even though sometimes it becomes a technology constraint to manage transactional data history of more than a few months. This paper presents a framework to efficiently manage massively large and complex datasets. The framework has been tested on a communication service provider producing massively large complex streaming data in binary format. The communication industry is bound by the regulators to manage history of their subscribers' call records where every action of a subscriber generates a record. Also, managing and analyzing transactional data allows service providers to better understand their customers' behavior, for example, deep packet inspection requires transactional internet usage data to explain internet usage behaviour of the subscribers. However, current relational database systems limit service providers to only maintain history at semantic level which is aggregated at subscriber level. The framework addresses these challenges by leveraging Big Data technology which optimally manages and allows deep analysis of complex datasets. The framework has been applied to offload existing Intelligent Network Mediation and relational Data Warehouse of the service provider on Big Data. The service provider has 50+ million subscriber-base with yearly growth of 7-10%. The end-to-end process takes not more than 10 minutes which involves binary to ASCII decoding of call detail records, stitching of all the interrogations against a call (transformations) and aggregations of all the call records of a subscriber.

Keywords : big data, communication service providers, enterprise data warehouse, stream computing, Telco IN Mediation

Conference Title : ICBDC 2018 : International Conference on Big Data Computing

Conference Location : Dublin, Ireland

Conference Dates : February 15-16, 2018